Computer Networks - Midterm

Prof. J.-P. Hubaux and Dr. M. H. Manshaei

November 9, 2010

Duration: 1:45 hours, closed book.

Please write your answers on these sheets in a ***readable*** way.
Poorly written answers will ***not*** be corrected.
Use extra sheets if necessary (put your name on them).

You may write your answers in English or in French.

The total number of points is 40.

This document contains 14 pages.

First Name (Prénom):
Last Name (Nom de famille):

SCIPER No:

Division: ☐ Communication Systems ☐ Computer Science
☐ Other (mention it): . . . .

Year: ☐ Bachelor Year 2 ☐ Bachelor Year 3
☐ Other (mention it): . . . .

(answers to the questions are shown in italic and blue)

1 Short questions (5 points)

For each question, please circle a single best answer.

1. Statistical multiplexing is:
   (a) a mechanism providing circuit-like behavior in packet-switched networks
   (b) a method to optimally share files in a peer-to-peer system
   (c) an algorithm that divides up the frequency spectrum of a link
   (d) an on-demand form of sharing a transmission link *CORRECT*

2. The end-to-end delay depends on:
   (a) the congestion of the network
   (b) the number of hops between source and destination
   (c) the application-layer protocols
   (d) the processing delay in the routers
   (e) all of the above
   (f) (a), (b) and (d) *CORRECT*

3. A tier-2 ISP:
   (a) is a provider of one or more tier-1 ISPs
   (b) is connected to only one tier-1 ISP
   (c) can be connected to other tier-2 ISPs *CORRECT*
   (d) Both (b) and (c)

4. HTTP with persistent connections
   (a) requires 2 RTTs per object
   (b) requires the server to open a new TCP connection to send a response
   (c) provides higher security against Denial of Service (DoS) attacks
   (d) none of the above *CORRECT*

5. Cookies enable
   (a) a Web server to infect a user's machine with malware
   (b) a Web server to track a user's activity at its own Web site *CORRECT*
   (c) a Web server to know all previous Web pages visited by the user
   (d) a Web server to learn your full name, e-mail address and credit card information

6. IMAP
   (a) is built on top of POP3
   (b) is stateless across sessions
   (c) allows users to create remote folders *CORRECT*
   (d) is not compatible with SMTP

7. In a reliable data transfer protocol, duplicate ACK
   (a) is a technique to increase the probability of successful ACK transmission.
   (b) is used to decrease the congestion window at the sender.
   (c) can be used instead of NAK packets.
   (d) is used to detect packet loss before the timeout.
   (e) both (c) and (d) *CORRECT*

8. By replacing a stop-and-wait protocol with a pipeline protocol, a sender increases its link utilization by a factor of k. How many packets does the sender allow to transmit without receiving acknowledgements?
   (a) $k/2$
   (b) k *CORRECT*
   (c) $\sqrt{k}$
   (d) $2k$

9. In a TCP connection, the timeout value is 1 second. What will be the new value of the timeout, when the timer expires?
   (a) 1.5 second
   (b) 2 seconds *CORRECT*
   (c) It depends on *SampleRTT* and *EstimatedRTT* values.
   (d) $1 + \alpha \cdot SampleRTT$, where $\alpha = 0.125$.

10. Consider the following Java code:
    `socket = new ServerSocket();`
    After executing this code, the `socket` object will be bound/connected to:
    (a) A local IP address.
    (b) A local IP address and port number. *CORRECT*
    (c) Two IP addresses: local and remote.
    (d) Two IP addresses and ports: local and remote.

2 Web and HTTP (10 points)

A student is performing a search query on the Google Search Engine webpage. The generated traffic was captured with Wireshark and the traces are given in Figure 1, Figure 2, Figure 3 and Figure 4. By analyzing the given traces, answer the following questions:

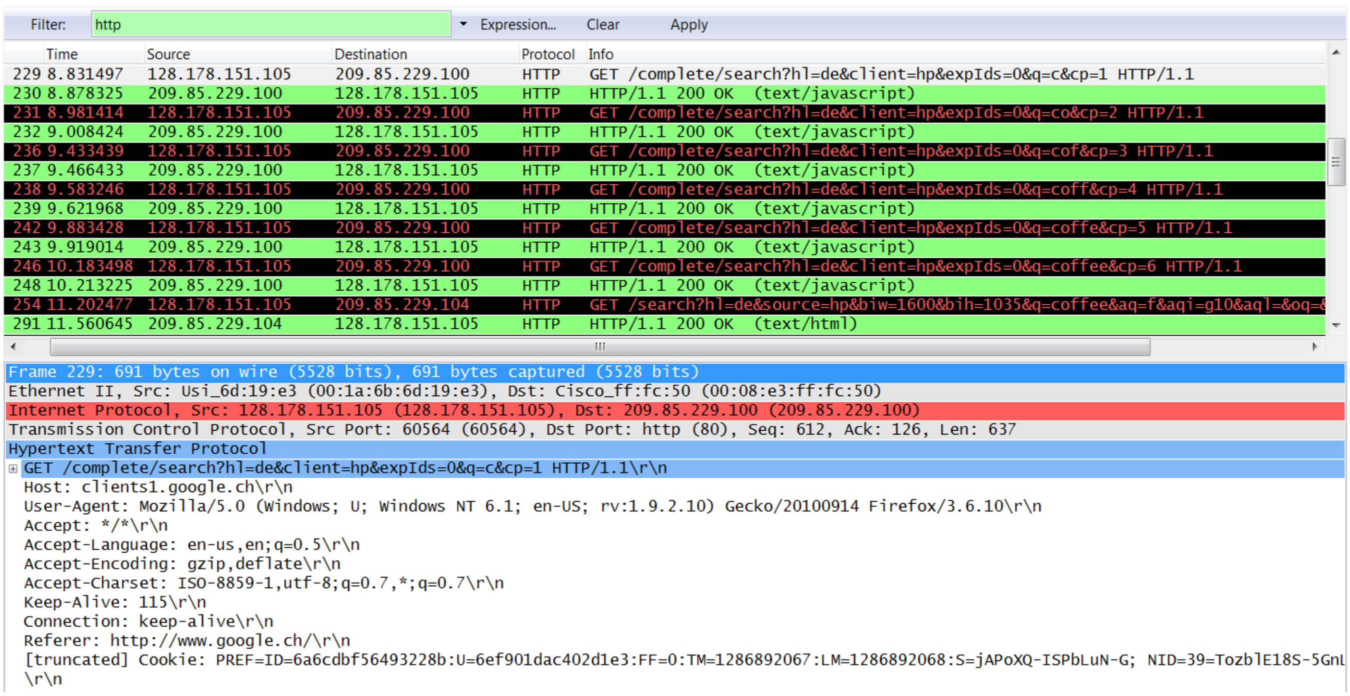

Figure 1: Wireshark traces - The first HTTP GET request

Question 1: What is the *URL* of the webpage on which the student types the keyword?

The URL is www.google.ch.

Question 2: How many HTTP GET requests are sent from the student's machine?

7 HTTP GET requests.

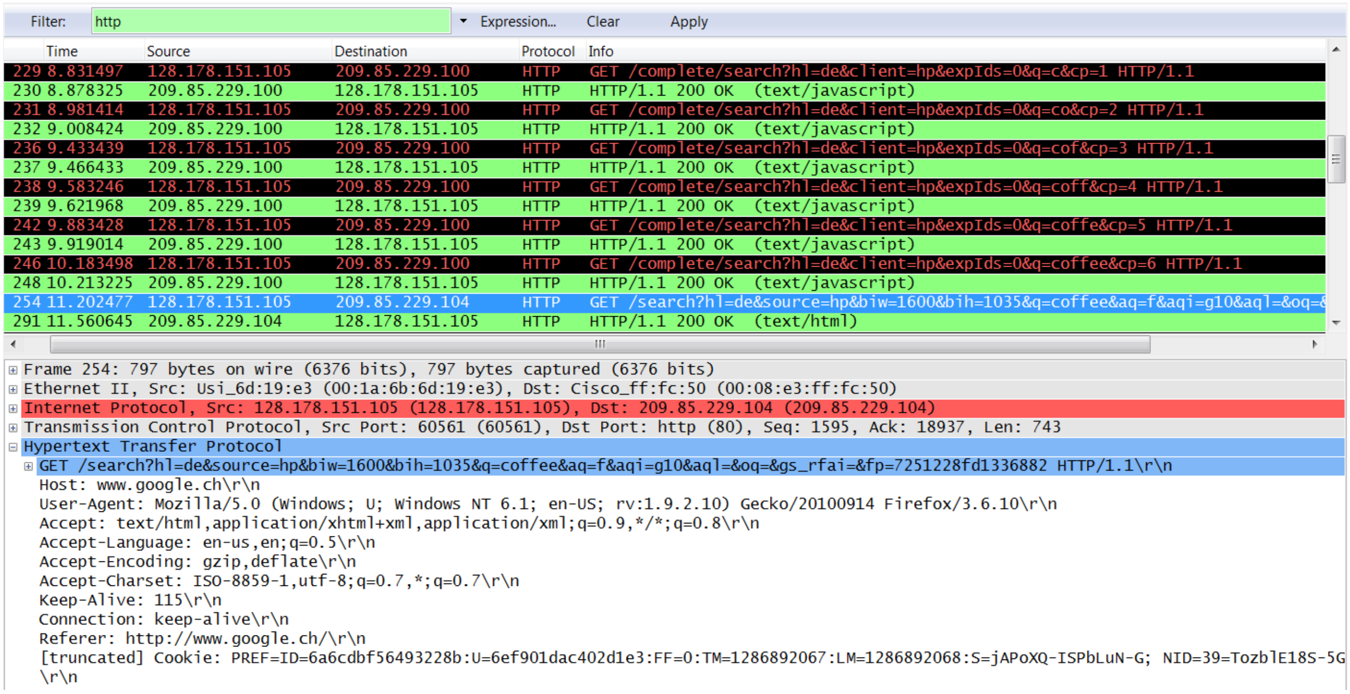

Figure 2: Wireshark traces - The last HTTP GET request

Question 3: Are all of the HTTP GET requests sent to the same host? Write the name(s) and IP address(es) of the host(s).

The first six HTTP GET requests are sent to the host clients1.google.ch with IP address 209.85.229.100. The last HTTP GET request is sent to host www.google.ch with IP address 209.85.229.104.

Question 4: What is the search query that student has entered?

The keyword is: ***coffee****.*

Question 5: Why are there more than one HTTP GET requests?

As you type, Google's algorithm predicts and displays search queries based on other users' search activities. These searches are algorithmically determined based on a number of purely objective factors (including popularity of search terms) without human intervention. All of the predicted queries shown have been typed previously by other Google users. Therefore, after each of the letters of the search query is entered, an HTTP GET request is generated by the browser.

Question 6: What does each of the HTTP GET requests correspond to?

The first six HTTP GET requests correspond to autocomplete functionality of Google Search Engine with the following search queries: "c", "co", "cof", "coff", "coffe" and "coffee", respectively. They are generated by the student typing each additional letter of the search query. The last HTTP GET is a search query with a complete keyword "coffee".

Question 7: How many TCP connections are opened? Explain your answer and relate it to the given traces.

Based on the trace of the first HTTP GET request (Figure 1), the browser opens a persistent TCP connection with the server. Therefore, all the subsequent content requested and sent from this server (corresponding to the responses to the HTTP GET requests 2-6) will be downloaded via the same TCP connection. You can conclude this based on the value of the "Connection" HTTP header ("Connection: keep-alive") in the first HTTP GET request. The last HTTP GET request (Figure 2) is sent to a different host, with a different IP address, therefore that content will be downloaded via a different TCP connection (which is also a persistent TCP connection based on the HTTP header "Connection: keep-alive").

Question 8: Is the first HTTP GET request (Figure 1) the first request ever that is sent from the student's machine to Google servers? Explain your answer and relate it to the given traces.

> *No, because a cookie is associated with the first HTTP GET request, thus the student has already visited this domain in the past.*

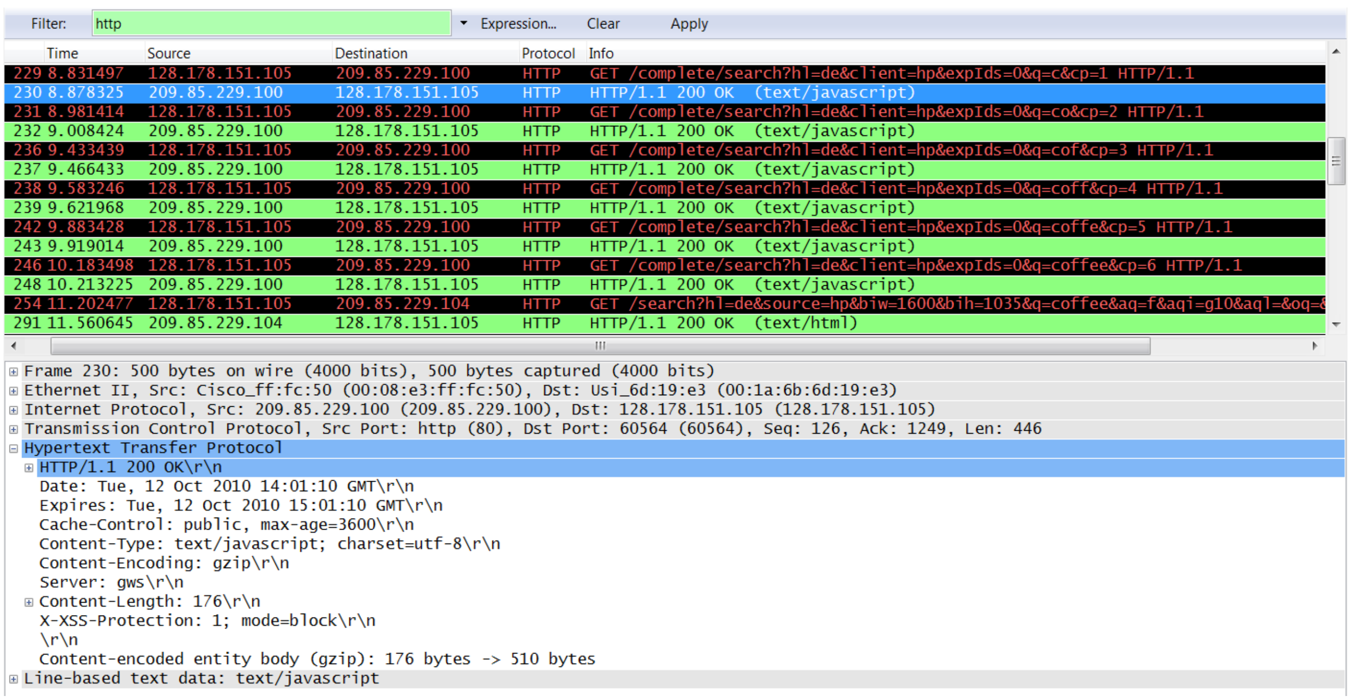

Figure 3: Wireshark traces - The first HTTP response from the web server

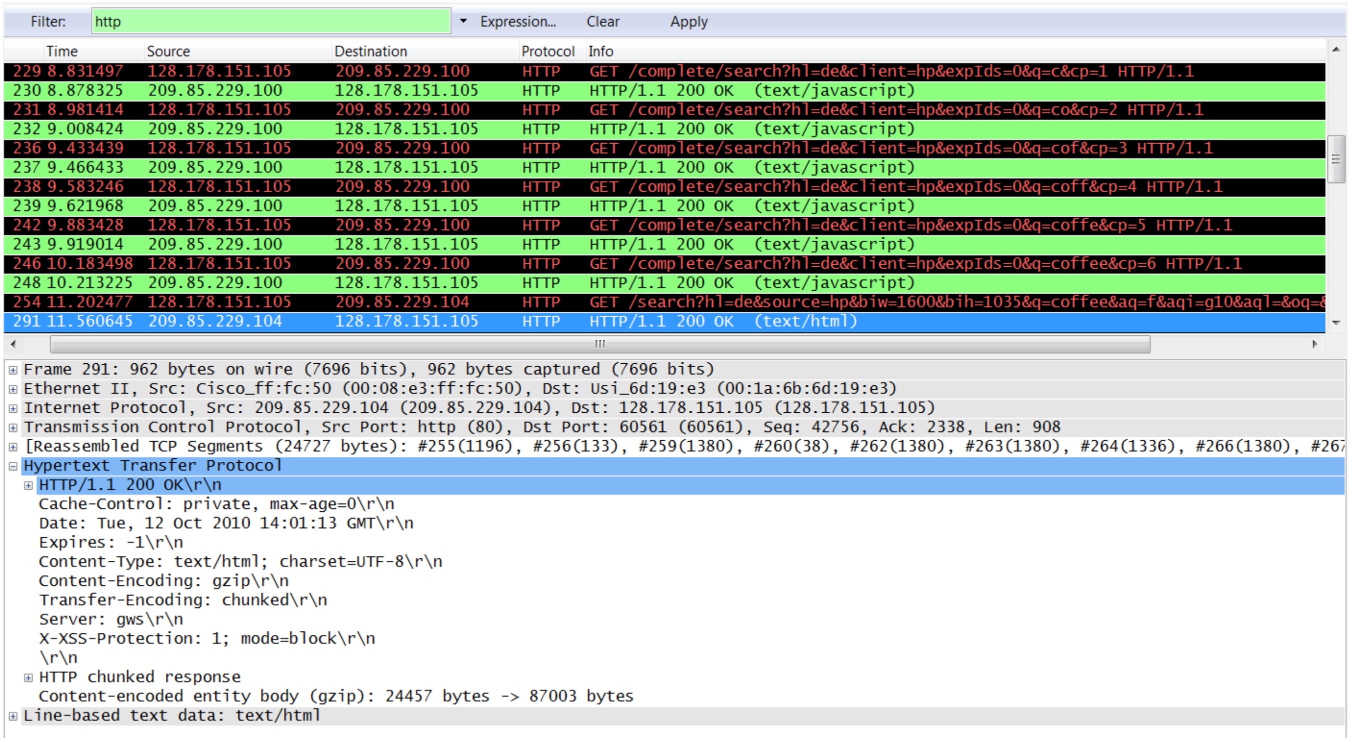

Figure 4: Wireshark traces - The last HTTP response from the web server

Question 9: Can the content of the HTTP responses to each of the HTTP GET requests be cached at the EPFL web cache? Explain your answer and relate it to the given traces.

The content of the responses corresponding to the first six HTTP GET requests (autocomplete results) can be cached. You can conclude this based on the value of the "Cache-control" HTTP header ("Cache-Control: public, max-age=3600") in the response from the server (Figure 3). The content of the response corresponding to the last HTTP GET request (search result) cannot be cached. You can conclude this based on the value of the "Cache-control" HTTP header ("Cache-control: private, max-age=0") in the response from the server (Figure 4).

3 Network Delays (8 points)

Question 1: Calculate the total time required to transfer a 1.5 MB file in the following cases, assuming a RTT of 80 ms, a packet size of 1 KB data, and an initial 2·RTT of "handshaking" before data is sent. Neglect the overhead due to headers. (Note: 1 MB = 1024 KB, 1 KB = 1024 B, 1 Mbps = 10^6 bps)

a. The transmission rate is 10 Mbps, and data packets can be sent continuously.

We consider the transfer as completed when the last data bit arrives at its destination.
1.5 MB = 12582912 bits.
2 initial RTTs (160 ms) $+\frac{12582912}{10000000bps}$ *(transmit)* $+RTT/2$ *(propagation)* ≈ 1.458 *seconds.*

b. The transmission rate is 10 Mbps, but after the source finishes sending each data packet it must wait one RTT before sending the next.

Number of packets required = 1.5 MB / 1 KB = 1536. To the computation done in (a) we add the time for 1535 RTTs (the number of RTTs between when the packet 1 arrives and packet 1536 arrives), for a total of $1.458 + 122.8 = 124.258$ *seconds.*

c. The link allows infinitely fast transmit, but limits bandwidth such that only 20 packets can be sent per RTT.

Dividing the 1536 packets by 20 gives 76.8. This will take 76.5 RTTs (half an RTT for the first batch to arrive, plus 76 RTTs between the first batch and the 77th partial batch), plus the initial 2 RTTs, for 6.28 seconds.

d. The link allows infinitely fast transmit, but during the first RTT the source can send one packet (2^{1-1}), during the second RTT it can send two packets (2^{2-1}), during the third it can send four packets (2^{3-1}), and so on.

Right after the handshaking is done we send one packet. One RTT after the handshake the source sends two packets. At n RTTs past the initial handshaking we have sent $1 + 2 + 4 + ... + 2^n = 2^{n+1} - 1$ packets. At $n = 10$ the source bas been able to send all 1536 packets; the last batch arrives 0.5 RTT later. Total time is $(2 + 10.5)$ RTTs, or 1 second.

4 Peer-to-peer (5 points)

Question 1: Consider the following simplified BitTorrent scenario. There is a swarm of 2^n peers and, during a considered time period, no peers join or leave the swarm. It takes a peer 1 unit of time to upload or download a piece, during which time it can only do one or the other. Initially one peer has the whole file and the others have nothing.

a. If the swarms target file consists of only 1 piece, what is the minimum time necessary for all the peers to obtain the file? Consider only the upload/download time and ignore everything else.

All peers could have the file after n time units. During each time unit, each peer with the piece can transmit it to one peer without the piece, so the number of peers with the piece doubles with each time unit: 1 ($= 2^0$) at time 0, 2 ($= 2^1$) at time 1, 4 ($= 2^2$) at time 2, up to 2^n peers at time n.

b. Let x be your answer to the preceding question. If the swarms target file instead consisted of 2 pieces, would it be possible for all the peers to obtain the file in less than $2x$ time units? Why or why not?

All peers could have the file after less than $2n$ time units. If all pieces were downloaded to just the right peers in just the right order, it would be possible for all peers to obtain the file after as few as $n + 2$ time units. Let's label the two pieces A and B. Let's label as PA the peer that initially has the file. During the first time unit, PA transmits B to another peer, call that other peer PB. Split the peers into two equal groups of $2(n - 1)$, one containing PA and the other containing PB. Now, from the result of the first question, we know that all the peers grouped with PA can obtain A within an additional $n - 1$ time units. Because the two sets of peers are disjoint, with no interference or contention between them, all the peers grouped with PB can obtain B during the same $n - 1$ time units. Together with the initial step, we have used n time units so far. Another time unit will suffice for the peers grouped with PA to transmit A to all the peers grouped with PB. One more time unit will suffice for the peers grouped with PB to transmit B to all the peers grouped with PA (except PA itself, which has had B from the beginning). The 2 time units required for each half to transmit its piece to the other half increases the total to $n + 2$ time units.

5 Transport Layer (8 points)

Assume that host A wants to send 4 segments (of L Bytes each). The timeout values at hosts A and B are $2 \cdot RTT$. Suppose that the 2^{nd} segment is lost whereas all the following segments and all acknowledgments are delivered without error. The hosts can use Go-Back-N (GBN), Selective Repeat (SR), or TCP at the transport layer. The window size in all protocols is greater than 10.

Question 1: Draw the missing segments and acknowledgments with their sequence numbers in Figure 5, for all three protocols (You do not need to consider delayed ACK in TCP.).

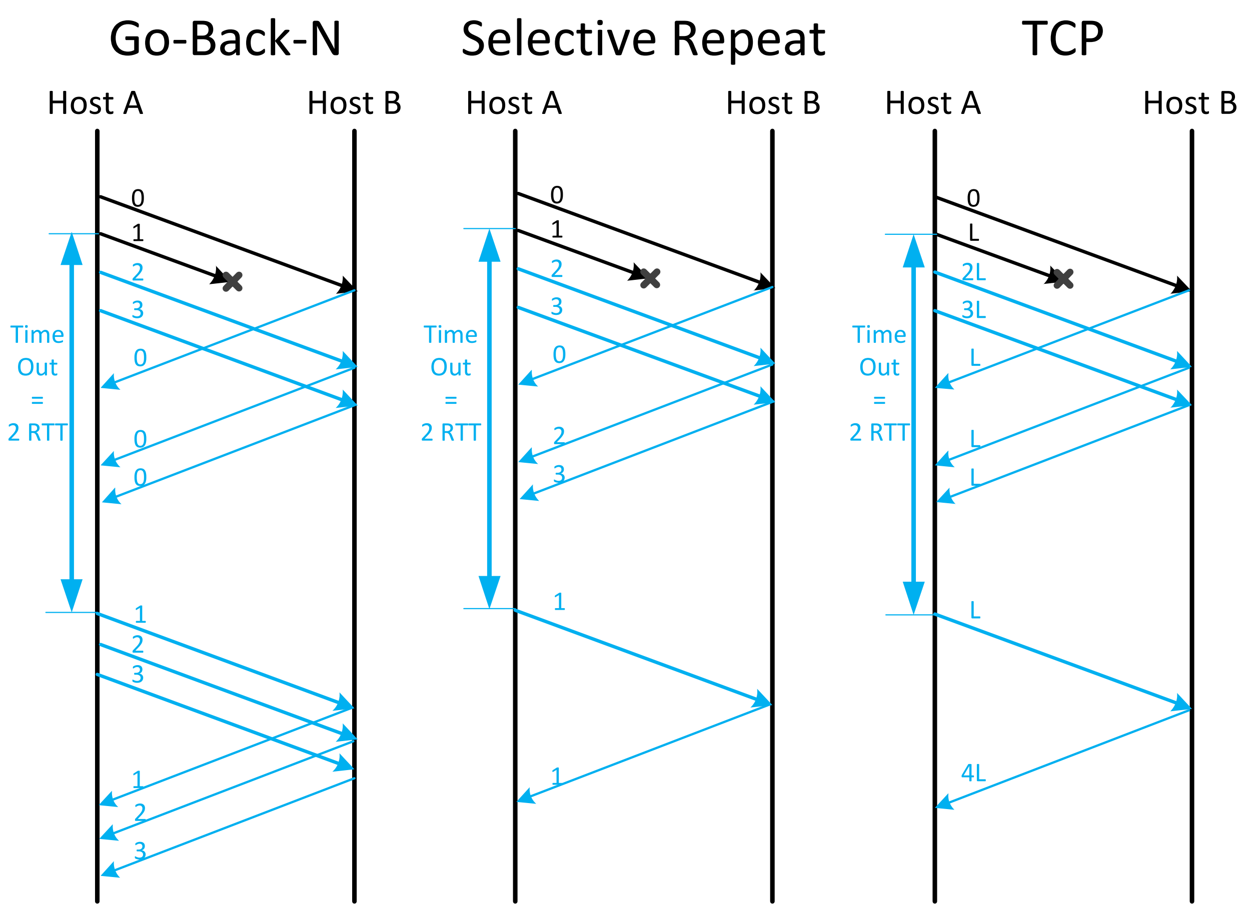

Figure 5: Go-Back-N, SR, and TCP operation with 4 packets.

Question 2: Compare the throughput of host A achieved with the above protocols. Which protocol obtains the maximum throughput? Why?

Considering the solution presented in Figure 5, the throughput is almost the same with all protocols and it is equal to $\frac{4L}{3RTT}$. *Note that the throughput of TCP and SR is slightly greater than the GBN protocol.*

Question 3: Assume that Host A has 5 segments to transmit and the 2^{nd} segment is lost, but all the following segments and all acknowledgments are delivered without error. For each protocol, calculate the throughput of host A and compare with question 2. Which protocol obtains the maximum throughput with 5 segments? Justify your answer.

The throughput of Host A *will increase to* $\frac{5L}{3RTT}$*, by using GBN and SR. TCP obtains the maximum throughput as it uses the duplicate ACKs to detect the packet loss event and consequently resends packet* 1 *after receiving* 3 *ACKs for packet* 1.

6 TCP- Congestion Control (4 points)

In the congestion avoidance phase of TCP, the sender probes for additional bandwidth by increasing congestion window (cwnd) by 1 MSS each RTT, until a loss event occurs. In the macroscopic description, we assume that W is the value of congestion window when a loss event occurs. We also assume that the RTT and W are approximately constant over the duration of the connection. Let us assume that the sender increases the cwnd by x MSS each RTT ($x > 1$), during the congestion avoidance phase.

Question 1: Calculate the new packet loss rate of the sender by using the macroscopic description of TCP. Will the sender be able to decrease the packet loss rate, by using this technique? Justify your answer.

Hint 1: We assume that $W/(2x)$ is an integer.
Hint 2: $\sum_{i=0}^{b} a \cdot i = \frac{ab}{2}(1+b)$.

The loss rate, L, is the ratio of the number of packets lost over the number of packets sent. In a cycle, 1 packet is lost. The number of packets sent in a cycle can be calculated as:

$$\frac{W}{2} + (\frac{W}{2} + 1) + \cdots + W = \frac{W}{2}(1 + \frac{W}{2}) + \sum_{i=0}^{W/2} i = \frac{W}{2}(1 + \frac{W}{2}) + \frac{W}{4} + \frac{W^2}{8} = \frac{3}{8}W^2 + \frac{3}{4}W.$$

Now if the sender sends x MSS instead of 1 the number of transmitted packets is:

$$\frac{W}{2} + (\frac{W}{2} + x) + (\frac{W}{2} + 2x) + \cdots + (\frac{W}{2} + \frac{W}{2x}x) = (\frac{W}{2x} + 1)\frac{W}{2} + \sum_{i=0}^{W/(2x)} ix = (\frac{W}{2x} + 1)\frac{W}{2} + \frac{W}{4}(1 + \frac{W}{2x}) = \frac{3}{8x}W^2 + \frac{3}{4}W.$$

This shows that the number of transmitted packets decreases and consequently the packet loss rate increases. Hence she/he is not able to decrease the packet loss rate.